



Opinion mining from machine translated Bangla reviews with stacked contractive auto-encoders

Matteo Bodini¹

Received: 31 March 2021 / Accepted: 9 February 2022 / Published online: 25 February 2022
© The Author(s) 2022

Abstract

In the last years, online users have been sharing more and more opinions, reviews, and comments on the web. Opinion mining is the automatic process of getting the subject of such opinions, and recently it has been attracting great commercial and academic interest. Several methods were presented for performing opinion mining in Bangla language, however they reported limited performance. In the present article, we considered the only two publicly datasets available for opinion mining in the Bangla language. We machine translated the datasets into the English language and we preprocessed them by extracting textual frequency based features. Then, we designed two stacked contractive auto-encoders based architectures to perform opinion mining in Bangla language, one for each dataset. The classifiers were trained on the machine translated version on the two datasets in a stacked learning fashion. The proposed classifiers achieved improved performance, with respect to accuracy ($\geq 96\%$), precision ($\geq 93\%$), recall ($\geq 94\%$), and F1 score ($\geq 94\%$), reported in the past state of the art works. Furthermore, the experimental results showed that both the machine translation procedure and the stacked learning frameworks improved the final classification performance.

Keywords Text classification · Aspect-based sentiment analysis · Opinion mining · Bangla language · Auto-encoders · Stacked learning · Machine translation

1 Introduction

Latest research works showed that Internet users sometimes trust more to online reviews than their relatives or friends (Lăzăroiu et al. 2020). From companies point of view, such opinions tremendously influence the purchase of their services and products. Indeed, in order to prevent people from understanding the main drawbacks of their products and services, companies often want to limit online users' participation in reviewing them. This behavior represents a crucial point of many big companies' market strategy in a more and more competitive world (Trusov et al. 2009).

Sentiment Analysis (SA) studies how to determine the viewpoint of people on a certain topic (Yue et al. 2019; Li et al. 2019; Hacid et al. 2018). Usually, the most tackled task within the context of SA is the classification the polarity of

documents, where a document can represent a news article, a social network post on Twitter, a review of a restaurant on Tripadvisor®, etc. The polarity stands for the valence with which the ideas of the user are conveyed. Usually it considered as a discrete value and it could be negative, neutral or positive.

SA can be carried at three main levels (Yue et al. 2019; Li et al. 2019; Hacid et al. 2018): at the document level SA classifies the whole opinion expressed in the document with positive, negative or neutral sentiment. At the sentence level SA classifies if each sentence conveys a positive, negative or neutral sentiment. This kind of analysis is often performed on comments and reviews composed of one sentence. The third level is the aspect level, usually referred as opinion mining or Aspect-Based Sentiment Analysis (ABSA). In ABSA it is assumed that the document focuses on different aspects and it expresses different polarities for each of them. The aspects can be clearly recognized in the text or being implicitly present in it. Thus the task of ABSA is twofold: it first requires the classification of aspects on which the document is focused, and this sub-task is usually identified as opinion mining or

✉ Matteo Bodini
matteo.bodini@unimi.it

¹ Dipartimento di Economia, Management e Metodi Quantitativi, Università degli Studi di Milano, Via Conservatorio 7, 20122 Milan, Italy

aspect classification/extraction. Then, it follows the classification of the polarity of the recognized aspects.

To clarify the concepts of aspect and polarity, the reader may consider the following simplified version of a restaurant review:

“Amazing service and delicious food.”

There are two aspects in the above example: service and food expressed with positive polarity. Aspects can be obviously identified in the above example since the words “services” and “food” are clearly present in the text. However, sentences usually convey implicit aspects. For instance in the following example:

“Waiters were so nice and pasta was great.”,

the aspects “service” and “food” are still present, even if not clearly mentioned in the text.

Although researchers have been carrying extensive research on ABSA in the English language and in other languages (Yue et al. 2019), less research works have been published in the Bangla language where there is still the need to experiment new techniques, yet applied in other machine learning fields (Roy et al. 2013, 2017; Bodini et al. 2018; Boccignone et al. 2018), and even the need of datasets both labeled with multiple aspects and polarities (Ahmed et al. 2021). The most of the available datasets are labeled only with polarities and they are composed of a few thousand samples. Furthermore, several datasets are not provided in the original Bangla language and they are constructed by translating datasets from the English language, or from other ones.

In the last years, the need for ABSA in the Bangla language motivated researchers to create datasets for aspect and polarity classification in Bangla language. Recently, in Rahman and Dey (2018b) the authors presented two databases for benchmarking ABSA algorithms in Bangla language. The authors showed baseline results for the problem of aspect classification relying on classical machine learning algorithms. A Convolutional Neural Network (CNN) to classify aspects is presented in a following paper from the same authors, which gained improved performance (Rahman and Dey 2018a). Recently, a preliminary work from Bodini (2019a) introduced an auto-encoder based model that obtained promising performance on the dataset presented in Rahman and Dey (2018b). However, there is still the need of facing the problem of ABSA in Bangla language relying on low dimensional datasets and several machine learning methods still remains unexplored and then could be tested.

In the present article, it is faced the problem of opinion mining in the Bangla language. The most relevant findings, and contributions are as follows:

- Contractive Auto-Encoders (CAEs) (Rifai et al. 2011) trained in a stacked fashion were designed to perform opinion mining in the Bangla language. CAEs are a modified version of the standard auto-encoder (AE) models that were found to be effective in preventing the overfitting problem when working with low dimensional datasets. Further, standard AEs previously showed promising performance on the same problem (Bodini 2019a).
- The presented CAEs were trained in stacked fashion. Stacked learning, introduced in Vincent et al. (2008) has been proven to be an effective tool for simplifying the design and for improving the performance of AE models. A comparison with the standard end-to-end learning paradigm was provided.
- The latter advances in the field of machine translation has proven to be a concrete possibility for improving ABSA in less represented languages (Wan 2008; Balahur and Turchi 2014). The used datasets were machine translated into the English language and CAEs were trained on the translated datasets and on the original ones to provide a comparison.
- The proposed CAEs showed better numerical performance, with respect to the previous works that tackled the problem (Rahman and Dey (2018a, 2018b); Sazed and Sampath 2019; Bodini 2019a). The framework of stacked learning lead to an improvement in performance with respect to the standard end-to-end learning procedure. Further, CAEs performed better on the machine-translated version of the dataset, showing that machine translation improved performance.

The article is presented as follows: in the Sect. 2, we report the related articles that faced ABSA and SA in the Bangla language; in the Sect. 3, we report the description of the two datasets, namely the Cricket and Restaurant datasets, used to assess the performance of the proposed architectures; in the Sect. 4, we report the performed preprocessing and feature extraction, that is represented by the machine translation of the sentences contained in the datasets and the extraction of textual frequency based features; in the Sect. 5, we describe the CAE models that were used to build the two presented classifiers; in the Sect. 6, we describe the experimental settings to train the proposed architectures according to the stacked framework; in the Sect. 7, we report the experimental results; in the Sect. 8, we report the discussions, and finally in the Sect. 9 we draw the conclusions of our work and possible future directions.

2 Related works

In this section, we report the articles that faced ABSA and SA in the Bangla language. Even works that are focused on SA are reported since it is a strictly related task to ABSA and

several techniques and datasets are common between these two problems. More detailed and wide information is contained in the remarkable survey articles (Yue et al. 2019; Li et al. 2019). Concepts related to the field of machine learning can be deepened in the following books (Bishop 2006; Burkov 2019).

Considering the Bangla language, the majority of the literature only face SA. Chowdhury and Chowdhury (2014) automatically extract sentiments from Bangla blog posts and classify the polarity as positive or negative. A remarkable aspect of the article is that the authors exploited a semi-supervised bootstrap approach for constructing the training database, which makes unnecessary any manual annotation. For the classification step, the authors used Maximum Entropy and SVM (Support Vector Machine) and then they analyze the performance experimenting with many well-known text features.

Hasan et al. (2014) limits to SA and classify the polarity of phrases exploiting contextual variance analysis: in the field of linguistic, the valence of verbs indicates the number of close nouns with which such verb is related to. The system executes a step of parsing to catch the parts of the sentence in the first step and then it applies fixed rules to assign the contextual valence, i.e. the polarity, to the components of the sentence. Concerning databases for ABSA and SA in the Bangla language, most of them are not publicly available and they are kept private.

A valuable dataset for the task of SA is presented in Hassan et al. (2016). The database is named “Bangla and Romanized Bangla Texts” (BRBT). The database is composed of 9337 posts and it is the largest available in the Bangla language. The database contains also sentences written in Romanized Bangla, which is simply the Bangla language written with the Roman alphabet. However, the database is kept private at the moment (even if authors can share it by contacting them and filling a consent form). Hassan et al. designed a Long Short-Term Memory network (LSTM) and obtained high performance.

The above mentioned database is even used by Alam et al. (2017) where they use a CNN to tackle SA (considering only positive and negative polarity). The CNN model reaches 99.87% classification accuracy, which is 6.87% higher concerning (Hassan et al. 2016). We remark that the above-mentioned articles focus only on SA, as the underlying database is not annotated with aspects. Further, it is not simple to compare existing methods since the database is not public and the sharing is very limited.

In Rahman and Dey (2018b) baseline results for aspect extraction using k -Nearest Neighbor (k -NN), SVM, and Random Forests (RFs) are provided. The latter article is advanced by the work (Rahman and Dey 2018a) by the same two authors, where they classified aspects with a CNN and gained improved performance both for recall and F1-score.

The first input layer of the CNN is a phrase composed of concatenated word2vec word embeddings (Mikolov et al. 2013) and the CNN consists of only one convolutional layer, max pooling, and finally by a classification layer, composed of a fully connected network.

In Bodini (2019a) the authors proposed baseline AE models on the dataset presented in Rahman and Dey (2018b). The reported performance were improved with respect to the previous presented works mentioned in the above paragraph (Rahman and Dey 2018b, 2018). Even if the accuracy was not reported in the original article, the precision, recall and F1-score showed promising performance in the considered task, where the lowest value is represented by the recall in both the two datasets.

Regarding machine translation within the Bangla language, only the work of Sazzed and Sampath (2019) can be found. The authors present a bilingual approach to SA by comparing machine-translated Bangla corpus to its original form, and they worked on the Rahman et al. dataset (Rahman and Dey 2018b). They apply multiple machine learning algorithms: Logistic Regression, Ridge Regression, SVM, RF, Extra Randomized Trees, and Long Short-Term Memory. Their remarkable results of this work suggest that using machine translation improves classifiers' performance in both datasets.

The idea of using machine translation was even proposed by several studies that have been successfully conducted using cross-lingual methods. They can be roughly divided into two main categories (Abdalla and Hirst 2017): (1) methods that exploit parallel databases to train bilingual word embeddings (Tang et al. 2014); (2) methods that exploit bilingual lexicons, and machine translation systems (Zhou et al. 2016) to learn features from both languages. For instance, in Wan (2008), the authors designed a bilingual system to improve the performance of Chinese SA taking datasets and methods from English. To classify the polarity, the authors focused on unsupervised learning. In Balahur and Turchi (2014), the authors employed machine translation systems and supervised methods for multilingual SA. They focused on four languages: English, German, Spanish, and French. They used three machine translation systems Google, Bing, and Moses and they tested different supervised learning algorithms and various types of features. The deep discussions of the above works show that machine translation could be used for multilingual SA.

Despite the idea of employing machine translation has been successfully experimented, the novel research field of opinion mining make use of text preprocessing algorithms in order to analyze people's opinions, attitudes and emotions towards certain topics. However, it must be noted that sometimes part of such text preprocessing algorithms, which are mandatory for mining opinions on the Web 2.0, were shown to be sensitive to errors and mistakes contained in the

user generated content. In particular, while the most of the authors have taken grammatically false texts into consideration, the vast majority of research work assumes grammatically correct texts. But, as it is easy to incidentally notice, all social media channels contain many grammatical and orthographic errors which may decrease the final classification performance (Petz et al. 2015).

3 The cricket and restaurant datasets

In this section, The datasets published in Rahman and Dey (2018b) which were used to assess the performance of the proposed CAEs models are presented. The authors introduced two datasets named “Cricket” and “Restaurant”.

For the Cricket database the authors manually crawled 2900 opinions from two Facebook web pages: the Daily Prothom Alo and BBC Bangladesh, which are among the most popular news websites for the Bangla language. It is a shared guess that such websites publish authentic and trustworthy news and Bangla people frequently read news from these sources, even leaving comments for sharing their opinion. Either the Facebook webpages of Daily Prothom Alo and BBC Bangla count more than 10 million followers with thousands of posts.

The cricket topic was selected by the two authors because they report that Bangla people comment more on such sport concerning other topics, then it is more simple to get related posts (we remark that the crawling was done by hand). Here below, two opinions taken from BBC Bangladesh and Prothom Alo Facebook pages are reported. Original sentences are not easy to understand for non-native Bangla speakers, then translations in English language are reported.

For instance, the below one is a sentence related to the aspect “batting” and with negative polarity, taken from the BBC Bangla Facebook page:

“I don’t want to see Vijay in the national team anymore.”

The below one is a sentence with neutral polarity for the aspect “team”, taken from the Prothom Alo Facebook page:

“Razzaq is recently playing well, but the mister is not giving a chance to him.”

The Cricket database is jointly annotated by the two authors and several other people which include students, and colleagues from the University of Dhaka, Bangladesh. All the involved people categorized the entire database into five aspects: batting, bowling, team, team management and other. Three polarity levels were taken into account: positive, negative, and neutral polarities. Finally, the two authors used a majority voting method to take the final decision on the aspect and the polarity of a given comment.

For the Restaurant database the authors took inspiration from the English Restaurant database, introduced in SemEval-2014 by Pontiki et al. (2016): they manually translated the 2800 sentences into the Bangla language. Like in the original database, five aspects are present: food, price, service, ambiance, and miscellaneous. Further, the authors considered only three polarities: positive, negative, and neutral, despite in the original database also a polarity named “conflict” was present. The authors decided to consider conflict sentences as neutral sentences.

Again, here are reported two examples present in the database. The first represents aspect “food” with negative polarity:

“The food was good, but not enough.”

The second represents the aspect “ambiance” with negative polarity:

“There are a very limited number of seats.”

4 Preprocessing and feature extraction

In this section it is reported the preprocessing designed to compute features from the samples contained in the Cricket and restaurant datasets. The same preprocessing procedure was performed on both the Restaurant and Cricket databases.

First, Bangla sentences contained in the datasets were translated into the English language, using the APIs provided by Google machine translation online service. Next, all the noisy information that does not carry any semantic meaning in the considered contexts was removed from the sentences and replaced with the empty string. We removed stop words, punctuation, and numbers. This step is performed in most of the articles focused on ABSA and SA and researchers showed that this does not reduce the classification performance and aims in reducing the dimension of the feature space (Yue et al. 2019; Li et al. 2019).

Sentences were represented with the Vector Space Model (VSM), that is a model to represent documents as vectors of some specific identifiers (Salton et al. 1975). In the VSM model, a document d (that can be a sentence or even a text composed by several sentences) is represented by the vector $d = (w_1, w_2, \dots, w_t)$, where w_t is the weight w which is assigned to the t -th term present in the document d .

Several ways have been introduced to calculate the weights (Yue et al. 2019; Li et al. 2019). The Term Frequency-Inverse Document Frequency ($TF - IDF$) weighting was presented in Salton et al. Salton and Yu (1973) and it relies on the following two concepts: when the frequency of certain terms in a document is higher if compared to one of other terms, these terms can potentially better discriminate the document from documents within other categories.

When a term can be found with very high frequency all the available documents, it is unable to discriminate between the aspects for which documents are annotated (for instance, terms like “the”, “a”, and “an”, that are frequent in sentences). Thus, in $TF - IDF$ the term weights are proportional to their frequency, the Term Frequency - TF , while the specificity of terms can be calculated as the inverse of the documents where the terms are contained, The Inverse Document Frequency - IDF . The $TF - IDF$ measure is obtained as the product of the TF and IDF statistics and is defined as in in Eq. (1):

$$TF - IDF(t, d, D) = TF(t, d) \times IDF(t, D), \quad (1)$$

where t is a term contained in the document d and D is the set of all the documents. For a document $d \in D$ and a term $t \in d$, the weight w_t can be computed relying on Eq. (1) as it follows Eq. (2):

$$w_t = TF - IDF(t, d, D). \quad (2)$$

The most common definitions for both TF and IDF are reported in Eqs. (3,4) (Yue et al. 2019; Li et al. 2019; Schütze et al. 2008):

$$TF(t, d) = f_{t,d}, \quad (3)$$

$$IDF(t, D) = \log \frac{N}{n_t}, \quad (4)$$

where $f_{t,d}$ is the count for the term t in the document d , i.e., the times that the term t can be found inside the document d , $N = |D|$ and $n_t = |\{d \in D : t \in d\}|$, i.e., the documents where the term t can be found. Thus, the most common $TF - IDF$ measure is reported in Eq. (5):

$$TF - IDF(t, d, D) = f_{t,d} \times \log \frac{N}{n_t}. \quad (5)$$

The weights of the VSM representation were computed for each sentence relying on Eq. (5). Two common $TF - IDF$ measures were further selected and weights were computed even according to the following Eqs. (6,7) (Schütze et al. 2008):

$$TF - IDF'(t, d, D) = (1 + \log(f_{t,d})) \times \log \frac{N}{n_t}, \quad (6)$$

$$TF - IDF''(t, d, D) = \left(0.5 + 0.5 \frac{f_{t,d}}{\max_{\{t' \in d\}} f_{t',d}} \right) \times \log \frac{N}{n_t}. \quad (7)$$

5 Auto-encoder models

In the Sect. 5.1 it is presented the standard AE model. Then in the Sect. 5.2 it is presented the CAE model, which is a variant of the standard AE model used in the present work.

5.1 Standard auto-encoders

The AE models learn an identity function for which the output values should be similar to input ones according to a specific error measure, even called as loss measure or reconstruction error (Rumelhart et al. 1986). Within the learning process, AEs build an internal representation of the input of lower dimensionality and they are capable to recover the original input from such representation with the learned mapping. Despite they are designed for dimensionality reduction tasks, AEs are even used for binary and multi-class classification tasks (Hinton et al. 2006; Hinton and Salakhutdinov 2006; Bodini 2019b, c): AEs are first trained in an unsupervised fashion to learn the identity mapping, then the produced compact representation is fed into a classifier, e.g. k -NN, SVM or even a simple Softmax function (the reader can deepen the mentioned classifiers reading (Bishop 2006; Burkov 2019)).

A minimum AE architecture is composed of three levels: a input layer (L_1) of six neurons, a hidden layer (L_2) of four neurons, and an output layer (L_3) of six neurons. The nodes labeled with “+1” represent bias vectors, which are initially set as the unit vector. The input is the vector $\mathbf{x} = [x_1, x_2, \dots, x_6]$. The vector entries can stand for row input data or also features computed from it. The output is the estimate of the input $\hat{\mathbf{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_6]$. AEs learn the approximated identity function $\mathbf{x} \approx \hat{\mathbf{x}}$ to reconstruct outputs which are similar to inputs. Notice that the architecture of the AEs is usually symmetric, i.e. the number of neurons of the layer L_1 is the same of layer L_3 , and the same holds for architectures with multiple hidden layers.

The AEs work through encoding and decoding steps. The encoding is executed among the input and the next hidden layer (layers L_1 and L_2). In this step the training data is represented with the set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, where $\mathbf{x}_i \in R^{D_x}$, for $1 \leq i \leq m$, and D_x represent the dimension of the elements in the training set \mathbf{X} . The activation function of the hidden layer is usually a Sigmoid function, whose definition is reported in the Eq. (8):

$$\sigma(\mathbf{z}) = \frac{1}{1 + e^{-\mathbf{z}}}, \quad (8)$$

where $\mathbf{z} \in R^{D_h}$, and D_h is the dimension of the hidden layer. The Sigmoid function is applied elementwise to the elements of the input vector \mathbf{z} . Many other activation functions

may be employed. The output \mathbf{h} of neurons in the current hidden layer then is calculated as the Eq. (9):

$$\mathbf{h} = f(\mathbf{x}) = \sigma(\mathbf{W}\mathbf{x} + \mathbf{b}), \tag{9}$$

where $h \in R^{D_h}$, $\mathbf{W} \in R^{D_h \times D_x}$ is the matrix of the weights that associate the input layer and the hidden layer, and $\mathbf{b} \in R^{D_h}$ is a bias vector.

The decoding step is executed among the hidden and the next output layer (layers L_2 and L_3). The reconstructed copy of the input, $\mathbf{r} \in R^{D_x}$ is calculated as reported in the Eq. (10):

$$\mathbf{r} = g(\mathbf{x}) = \sigma(\mathbf{W}'\mathbf{h} + \mathbf{b}'), \tag{10}$$

where \mathbf{r} approximates the input \mathbf{x} , $\mathbf{W}' \in R^{D_x \times D_h}$ is the weight matrix and $\mathbf{b}' \in R^{D_x}$ is the bias vector.

During the training step, the two weight matrices \mathbf{W} , \mathbf{W}' and the bias vectors \mathbf{b} , \mathbf{b}' are updated to reach the minimum reconstruction error. For any definition of reconstruction error, it holds that the minimum error for an element \mathbf{x} is reached when $\mathbf{x} = \mathbf{r}$, where \mathbf{r} stands for the reconstructed copy of \mathbf{x} .

To evaluate the performance of an AE, the reconstruction error is usually computed over all the elements of the test set \mathbf{X}' as Eq. (11):

$$J_{AE}(\Theta) = \sum_{\mathbf{x} \in \mathbf{X}'} \mathcal{L}(\mathbf{x}, g(f(\mathbf{x}))), \tag{11}$$

where $\Theta = \{\mathbf{W}, \mathbf{b}, \mathbf{W}', \mathbf{b}'\}$ are the AE parameters, $\mathbf{X}' = \{\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m\}$ is the test set, and \mathcal{L} is the loss function. The latter is usually set as the binary cross-entropy when input values are in the range $[0, 1]$ or the mean squared error (MSE), which is reported in Eq. (12) Burkov (2019):

$$\mathcal{L}(\mathbf{x}, \mathbf{r}) = \mathcal{L}(\mathbf{x}, g(f(\mathbf{x}))) = \frac{1}{N} \sum_{i=1}^{D_x} [x_i - g(f(\mathbf{x}))_i]. \tag{12}$$

In the training step, the AE minimizes the loss function over the elements of the training set and then performance is evaluated on the test set computing Eq. (11).

5.2 Contractive auto encoders

In the recent years, several AEs have been introduced, where novelties were presented both in the architectures and in the design of loss functions. A recently proposed architecture is CAE, presented in Bengio et al. Rifai et al. (2011). The authors introduced a new loss where a regularizer is summed to the standard loss function of (11): it is added the Frobenius norm computed over the Jacobian of the input element \mathbf{x} . The norm is computed on the input element \mathbf{x} and it is the sum of squares of all the elements. The regularizer is reported as it follows Eq. (13):

$$\|J_f(\mathbf{x})\|_F^2 = \sum_{ij} \left(\frac{\partial h_j(\mathbf{x})}{\partial x_i} \right)^2. \tag{13}$$

When Eq. (13) is added to to the standard loss function of Eq. (11), the encoding becomes less sensitive to the small variations possibly present in training data. Therefore, it is potentially possible to reduce the sensitivity of the internal learned representation. The penalty term makes the internal space to not increase in dimension: maintaining a low dimension for the internal space makes CAEs more robust in the presence of noise in the input data, and further, it highly prevents overfitting phenomena.

The CAEs loss function over all the elements of the test set is as it follows in Eq. (14):

$$J_{CAE}(\Theta) = \sum_{\mathbf{x} \in \mathbf{X}'} [\mathcal{L}(\mathbf{x}, g(f(\mathbf{x}))) + \lambda \|J_f(\mathbf{x})\|_F^2], \tag{14}$$

where Θ and \mathbf{X}' are respectively parameters set and the test set. If \mathcal{L} is set as the MSE, the Eq (14) become as follows:

$$J_{CAE}(\Theta) = \frac{1}{N} \sum_{\mathbf{x} \in \mathbf{X}'} \sum_{i=1}^{D_x} [x_i - g(f(\mathbf{x}))_i + \lambda \|J_f(\mathbf{x})\|_F^2], \tag{15}$$

Bengio et al. showed that the proposed CAE obtain better results concerning many other AE architectures which are regularized with the weight decay method or for instance by denoising (Vincent et al. 2008): in particular, in their paper they show that the CAE is a better way than denoising autoencoders (DAEs) to extract reliable features on some well-known datasets (Rifai et al. 2011).

To compare with DAEs, CAEs explicitly force the robustness of the compact representation, whereas DAEs encourages robustness of reconstruction. DAEs' robustness is obtained in a stochastic way by having several explicitly corrupted versions of a training point aim for an identical reconstruction. On the other hand, CAEs' robustness to tiny perturbations is obtained analytically by penalizing the magnitude of first derivatives $\|J_f(\mathbf{x})\|_F^2$ at training points only, as shown in Eq. (14). In the case of a Sigmoid nonlinearity functions, the penalty on the Jacobian norm has the following expression of Eq. (16):

$$\|J_f(\mathbf{x})\|_F^2 = \sum_{i=1}^{D_h} [h_i(1 - h_i)]^2 \sum_{j=1}^{D_x} \mathbf{W}_{ij}^2 \tag{16}$$

Computing Eq. (16), has about the same cost as computing the reconstruction error and the overall computational complexity is $\mathcal{O}(D_x \times D_h)$ (Rifai et al. 2011).

6 Experimental settings

Features were extracted from sentences available in the Cricket and Restaurant datasets. A VSM vector representation was obtained for each sentence contained in the two datasets. Three separate tests were performed where the weights of the VSM vectors were computed in three different ways, as reported respectively in Eqs. (5, 6, 7).

The size of VSM vectors are different across the two datasets, as they are composed by sentences of variable length. To get a the same VSM length it was performed a zero-padding on the VSM representations to the size of the two longest sentences available in the two datasets. The VSM representations of Restaurant dataset sentences were zero-padded to a size of 51 and VSM representations of Cricket dataset sentences were zero-padded to a size of 34.

Two different CAE models were designed with 51 and 34 neurons for the input and output layers respectively for the Restaurant and Cricked dataset. The input and output of the two CAE was respectively the VSM representation of the sentence and its reconstruction. The Eq. (15) was set as loss function. ReLU activation functions were set for each neuron, where ReLU is defined as a function that returns the positive part of its argument: $ReLU(x) = \max(0, x)$. Adam was set for adaptive learning rate (Kingma and Ba 2014).

The two CAE models were trained for 100 epochs (where an epoch defines the number of times that a learning algorithm process the entire training dataset). A grid-search was performed in the interval [1, 100] to set the optimal number of neurons the in hidden layers, and in the interval [1, 10] to set the optimal number of hidden layers. After grid-search, two hidden layers were set for which the dimension is respectively 30 and 15 neurons for the CAE designed for the Cricket dataset, while 20 and 10 neurons for the CAE designed for the Restaurant dataset.

The optimal parameters were set assessing the average cross-validation error with 10-fold stratified cross-validation (where the distribution of classes is retained within the folds) on the training set taken from a random 70/30 train/test split. Relying on 10-fold cross-validation, the original dataset was randomly sampled into 10 equal size subsamples. Of the 10 subsamples, a single subsample was retained as the validation data for testing the model, and the remaining 9 subsamples were used as training data. The cross-validation process was then repeated 10 times (i.e. 10 folds), with each of the 10 subsamples used exactly once as the validation data. The 10 results from the folds were averaged to produce a single estimation (Bishop 2006; Burkov 2019).

The training was performed in stacked mode. Each layer was separately trained according to the loss reported in

Eq. (15) to learn the representation of the preceding one. To train the Restaurant CAE the following steps were performed: the first hidden layer was trained optimizing the loss in a CAE with a hidden layer composed of 20 neurons, and input and output layers of 51 neurons. The second hidden layer was trained optimizing the loss in a CAE with a hidden layer with 10 neurons, and input and output layers of 20 neurons. The first and second steps were performed in a symmetric way for the decoding part of the CAE. The Cricket CAE was trained in the same manner.

Two separate tests were made to perform a comparison in performance with CAE: (1) the architectures were even trained according to the standard end-to-end learning procedure. (2) The architectures were trained without adding the penalty term of Eq. (13), building in fact a standard AE model.

After training the two CAE, the decoding part of each architecture was removed (i.e. the layers after the third hidden layer) and it was added a Softmax layer. The outputs of such neurons, say z_1, z_2, \dots, z_5 , is computed for the j -th neuron in the last layer with the Softmax activation function as the in Eq. (17):

$$s_j = \text{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^5 e^{z_k}}, \quad (17)$$

for $1 \leq j \leq 5$. The five neurons represent the probability of classification for the five available aspect classes, i.e. food, price, service, ambiance and miscellaneous for the Restaurant dataset; batting, bowling, team, team management and other for the Cricket dataset. The input and output of the two CAE classifiers is respectively the VSM representation of the sentence and the probability of being classified in one of the available classes.

The CAE was finally trained to classify aspects. Categorical cross entropy (CE) was used as a loss function, defined as follows in Eq. (18):

$$CE(y_i, s_j) = - \sum_{i=1}^c y_i \log(s_j), \quad (18)$$

where y_i is the ground truth class, and s_j is the probability of each class, which is the output of the Softmax function of Eq. (17). CAEs were trained for 100 epochs on a 50/20/30 train/validation/test split and Adam was used as adaptive learning rate.

The preprocessing was performed using Python 3.7. The CAEs were implemented and trained in TensorFlow 2 library on Python 3.7. The CAEs were trained relying on Amazon Web Services EC2 p3.8xlarge instances with 4 Nvidia Tesla V100 GPUs, 64 GB of RAM, and 32 Intel Xeon Skylake CPUs.

7 Experimental results

The confusion matrices of the best two CAE-based classifiers were reported respectively in the Figure 1, panel 2a and panel 2b. The confusion matrices report the best performance reached in the case VSM vectors were computed relying on Eq. (7), and in the case when CAEs were trained with the stacked learning framework. The highest performance were obtained for the classes “Team” and “Other”, in the case of the Cricket dataset, and for the classes “Service”

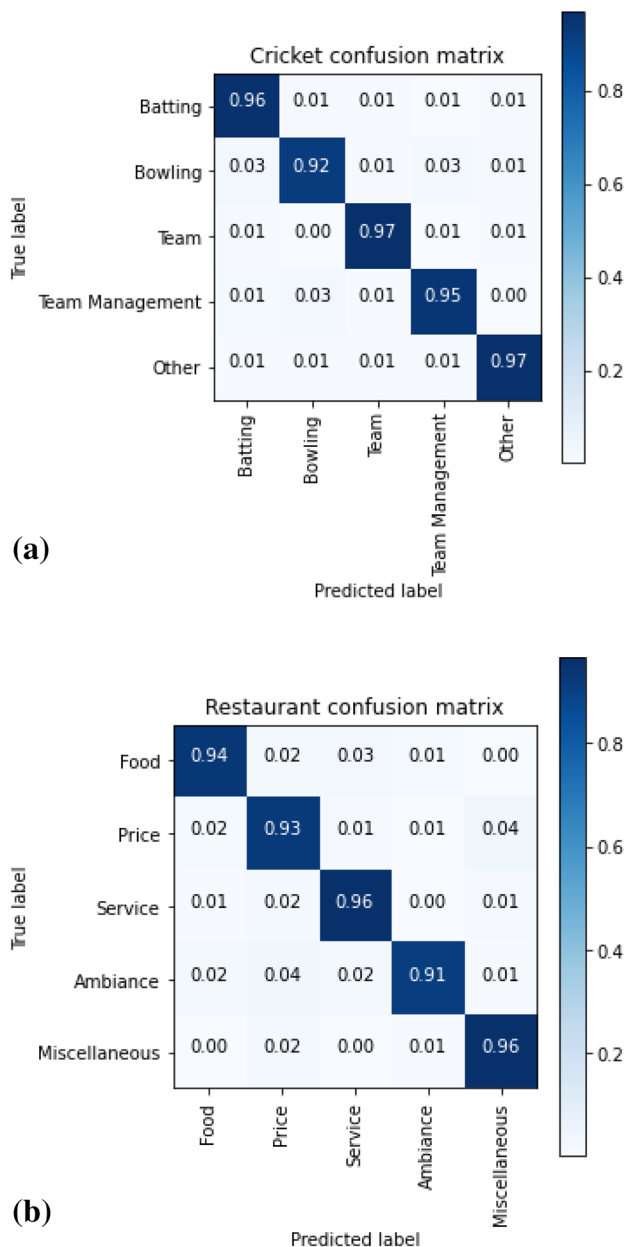


Fig. 1 The confusion matrices for the best performance CAE-based classifiers

and “Miscellaneous” in the case of Restaurant dataset. The lowest performance were obtained for the “Bowling” class in the Cricket dataset and for the “Ambiance” class in the Restaurant dataset.

Accuracy, precision, recall and F1-score were computed for each class relying on the confusion matrices reported in Fig. 1 and then averaged. Such measures were reported in the Table 1 for the two highest CAE-based classifiers and for other published works that relied on the same datasets. Accuracy was not reported in the case it was not available in the original publication.

In the Fig. 2 it was reported the evolution of the categorical cross entropy loss function and the accuracy measure with respect to the epochs respectively for the CAE-based classifiers on the Cricket dataset (panel 4a) and for the Restaurant dataset (panel 4b). The evolution was reported both on the training and validation set during the training phase.

In the Table 2 it is reported a comparison among the AE and CAE-based classifiers when training is performed in the standard end-to-end way and when stacked learning is adopted. The highest performance were obtained by the CAE-based classifiers for each metric, while the worse were obtained by the standard AE model trained in an end-to-end fashion.

8 Discussion

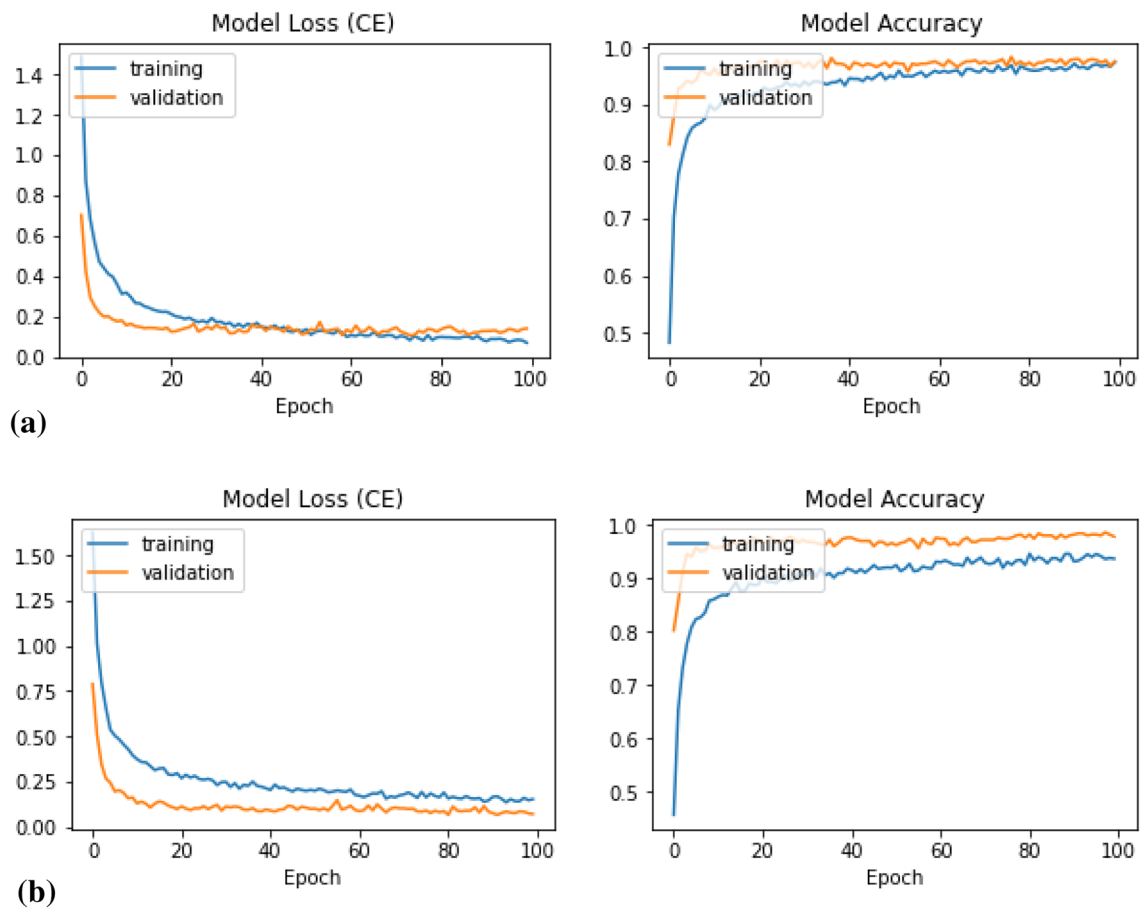
The confusion matrices reported in the Fig. 1 showed that the two proposed CAE-based classifiers reach high performance basically for each available class of the two datasets, even for the classes that are less represented, i.e. “Bowling” and “Team” (332 samples) for the Cricket dataset, and “Price” (178 samples) for the Restaurant dataset.

Even if the samples available are within the same order of magnitude for both the datasets, we reported in the Table 1 precision, recall and F1-score metrics to fairly evaluate the performance. The same choice was taken even by other researchers in the other works that relied on the same two datasets and sometimes accuracy it is not reported by them for such reason (Rahman and Dey 2018a, 2018b; Sazed and Sampath 2019).

In Rahman and Dey (2018b), SVM gained the better precision score respectively of 0.71 and 0.77 for the Cricket and Restaurant databases. Every baseline proposed method obtains very low recall and then F1-score and this mean that the proposed classifiers could be unable to handle the data imbalance problem. Our guess seems reasonable, but cannot be properly verified since the authors did not report the classification confusion matrices, but only averaged metrics. In Rahman and Dey (2018a), the proposed CNN obtain overall better performance if compared to baseline classifiers. Even if the precision is higher for the SVM, the

Table 1 Accuracy, precision, recall and F1-score for our models (in bold) and other works

Dataset	Model	Accuracy	Precision	Recall	F1-Score
Cricket	CNN (Rahman and Dey 2018a)	0.81	0.54	0.48	0.51
	SVM (Rahman and Dey 2018b)	/	0.71	0.22	0.34
	AE (Bodini 2019a)	/	0.93	0.85	0.88
	LSTM (Sazzed and Sampath 2019)	0.59	0.42	0.40	0.41
Restaurant	CNN (Rahman and Dey 2018a)	0.83	0.67	0.61	0.64
	<i>k</i> -NN (Rahman and Dey 2018b)	/	0.54	0.34	0.42
	AE (Bodini 2019a)	/	0.90	0.85	0.87
Cricket - translated	CAE	0.96	0.95	0.95	0.95
	LSTM (Sazzed and Sampath 2019)	0.72	0.41	0.44	0.43
Restaurant - translated	CAE	0.96	0.93	0.94	0.94

**Fig. 2** The evolution of categorical cross entropy loss and the accuracy on training and validation set for the Cricket (a) and Restaurant (b) CAE-based classifiers

CNN gain the most high recall and F1-score with a huge difference for both the databases: for the Cricket database it shows 0.51 F1-score, and for Restaurant database it shows 0.64 F1-score. The results suggest that the CNN better classifies aspects than baseline methods, and according to its recall score it returns could return a minor number of false positive classified items. The work of Sazzed and Sampath

(2019) shows improved performance in terms of accuracy with respect to Rahman et al., but worst in terms of precision, recall, and F1-score. Again, in the work it is not reported the confusion matrix of the LSTM network, but we guess that such classifier did not succeed in facing the unbalanced distribution of the dataset. In Bodini (2019a) the authors proposed baseline AE models whose performance

Table 2 Comparison of AEs and CAEs with respect to the training procedure

Dataset	Model	Accuracy	Precision	Recall	F1-Score
Cricket	AE	0.69	0.71	0.64	0.67
	CAE	0.75	0.89	0.77	0.82
	AE-stacked trained	0.84	0.79	0.70	0.74
	CAE-stacked trained	0.96	0.95	0.95	0.95
Restaurant	AE	0.63	0.73	0.67	0.69
	CAE	0.78	0.84	0.83	0.82
	AE-stacked trained	0.81	0.82	0.74	0.77
	CAE-stacked trained	0.96	0.93	0.94	0.94

suggested that AE could be valuable models in the ABSA task in the Bangla language. Even if the accuracy was not reported in the original article, the precision, recall and F1-score showed promising performance in the considered task, where the lowest value is represented by the recall in both the two datasets.

The reported performance in 1 shows that the proposed CAE-based classifiers lead to better performance with respect to the techniques presented in the previous published articles. Even if the datasets were unbalanced, we preferred to not balance them to compare with Rahman and Dey (2018a, 2018b) which did not make any use of any balancing technique. Further, we noticed that Sazzed and Sampath (2019) which made a class-balanced version of the dataset using the Synthetic Minority Over-Sampling Technique (SMOTE) algorithm did not improved the performance in a significant way. Differently from Rahman and Dey (2018a, 2018b), but like Sazzed and Sampath (2019), we performed an hyperparameter tuning strategy relying on 10 fold stratified cross-validation, that is a more robust robust estimate with respect to a 70/30 random split. Several training and test sets are used instead of a single training/test split, preventing the case in which the sampled sets are too difficult or too easy to classify. Further, the stratification sampling guarantees no class will be over-represented during the training procedure and avoid classifiers to be prone to overfit certain over represented classes. The value of 10 folds is frequently adopted and most of the applied machine learning studied that make use of k -fold cross validation rely on it, and it has been experimentally shown that 10 folds provide a fair estimate in terms of bias and variance (Kohavi 1995). Analyzing the Fig. 2 it is possible to conclude that the learning process is smooth and the presented model did not show any overfitting phenomena.

Among the three $TF - IDF$ used weighting methods, we obtained the best performance in the case of $TF - IDF''$,

reported in Eq. (7), for both the datasets. Other feature representation could be tested, where one is represented by the Bag of Words (BoW), used considered in Rahman and Dey (2018a). Differently from the VSM model, the specificity is not considered in BoW because usually it is built a dictionary where only the frequency count is considered. The $TF - IDF$ representation takes into account even the specificity with the IDF term, which could carry further information with respect to the BoW representation.

Looking again at the Table 1, it is clear from the work of Sazzed and Sampath (2019) that machine translation can help in improving the aspect classification performance. Even observing the previous and preliminary work of Bodini (2019a), it is possible to notice that machine translation lead to an improvement in classification performance over all the considered metrics. Without the need to be expert in the English language, it is evident that machine translation is not always accurate, since sometimes, lexical, grammar and semantic errors are present. The presence of misspelled words and differences in many regional words make word-to-word translations sometimes highly inaccurate. The complexity of Bangla language and the machine translation inability to relate words to the context, make it difficult to get the semantic meaning in some cases. However, it is remarkable to notice that not only the classification results was improved, but even that machine translation is capable of retaining sentiments and aspects in translated English sentences even though the translation itself is not precise, since it relies on an automatic service. Even by exploiting the possibly inaccurate machine translations, the proposed CAE-based classifiers performs better on the English translated dataset. A possible reason can be the reduced number of available words after the automatic translation, that dropped approximately the 19% of the previous words, by replacing them with synonyms or with simpler gramatic structures provided with the English language. It was not possible to deepen more this aspect due to the limited knowledge of Bangla language and this aspect should be more investigated in the future with the help of a native Bangla speaker.

In conclusion, it is even possible to see in Table 2 the performance improvements if CAEs are trained in the stacked learning scheme. As said in the preceding sections, stacked learning is a reliable training scheme that several times lead to improved performance if compared to the standard learning way. It learns better hidden embeddings by training the networks one level at time, with respect to train architectures which are initialized with random weights. A remarkable gain in performance is even present between AEs and CAEs. The results are reported without taking into account machine translation.

The CAE models were selected and experimented to face the problem of the *curse of dimensionality*. Recently, several methods have been proposed with the attempt of

both reducing the dimensionality and also improving feature extraction (Vincent et al. 2008; Bengio et al. 2006). Nowadays, with the latest advancements of the Web, a huge amount of data is made available and current works on SA and ABSA rely on feature spaces of huge dimensionality (Yousefpour et al. 2014). In the mentioned articles focused on ABSA in the Bangla language, the authors relied on several machine learning algorithms, for instance CNNs, SVM, k -NN, and RFs. Except for CNNs, handcrafted features well known in SA and ABSA research have been used. Authors used for instance Bag of Words (BoW) representation (Harris 1954), which can lead to very high dimensional representations when the number of words increases (Ivica and Zisserman 2009). Since in the field of ABSA researchers usually need to tackle the problem of dimensionality, a reduction of feature spaces is a wanted goal. This step was addressed many times in the available literature using dimensionality reduction techniques, such as principal component analysis, in the case researchers used the standard machine learning pipeline (composed of feature extraction, followed by a classification step). It was even addressed when researchers used CNNs, as the max pooling layers address the problem performing a subsampling, hence they reduce the dimensionality, but they did not reach high performance.

In the present work the dimensionality reduction problem was addressed differently from the previous work by exploiting CAEs. Despite such model have been poorly used to tackle the ABSA problem, they have been proven to be valid in retaining only the important features from very high dimensional spaces (Vincent et al. 2008; Bengio et al. 2006; Bodini 2019a). Once the CAEs were set up after the learning step, the classification problem was naturally tackled: the proper compact representation learned by CAEs was feed to a classifier, even a simple Softmax one, to predict the class of the sentences.

9 Conclusions

ASBA in the Bangla language is nowadays an important issue as the use of Internet is becoming more and more popular in Bangladesh: people use the internet in every moment of their life. Automatic classification of texts in the Bangla language could become crucial, as people could prefer to take decisions after reading the thoughts of other users.

Despite the advances of the research on SA in Bangla, the research on ABSA is limited because of the poor availability of the datasets. The used Cricket dataset consists of 2900 comments and the Restaurant dataset contains 2800 comments. Such low dimension could mean that with test datasets of higher dimension, the performance of the

proposed models may be significantly lower. For future works, we are planning to use web scraping frameworks in order to expand the results of the presented research to datasets with higher dimension, thus obtaining more significant performance. Furthermore, The performance of ABSA in Bangla is often negatively influenced by the complexity of the language, complex writing system and alphabet, few labeled data, and lack of methods and literature that address the problem.

The obtained results showed that stacked CAEs with machine translation can help in providing improved performance concerning the original language and can be exploited in the pipeline of ABSA and SA for Bangla and maybe other less spoken languages. We proved that, even though the Google machine translation system is not perfect in the Bengla-English translation, it can be experimented for ABSA in the Bangla language. Relying on the two available databases, we presented two CAE models to tackle the subtask of opinion mining, which deserves to be more developed. We compared the models with the other state-of-the-art approaches and the introduced CAEs gained better performance concerning previous models.

For the following works, we will experiment even polarity classification. In this way, we could compare with a significant number of works that are focused on SA in the Bangla language. This permits us to tackle the complete ABSA problem. One immediate way we plan to test in the next future is to use the compact representation we learned in a transfer learning scheme: performing a fine-tuning on the SA problem, using the same models, could lead to good performance. Another future direction is represented by the possibility of translating the available datasets and models from the English language to improve the performance on the Bangla language. There are several datasets designed for opinion mining in the English language that could be translated in the Bangla language to improve the opinion mining in the Bangla language. Finally, we will also tackle explainability: the majority of the proposed articles which tackle SA and ABSA don't analyze the explainability of the embeddings. In Weston et al. (2014) it is introduced a CNN which predicts tags for social network posts and even generates meaningful embeddings. These embeddings are proved to be meaningful because they have been also successfully tested on other problems. We plan to follow this article, to provide explainable classifiers, even with good performance.

Funding This research received no external funding.

Declarations

Conflicts of interest The author declares no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdalla M, Hirst G (2017) Cross-lingual sentiment analysis without (good) translation. In: Proceedings of the eighth international joint conference on natural language processing, Asian Federation of Natural Language Processing: Taipei, Taiwan, vol 1, pp 506–515
- Ahmed MBU, Podder AA, Chowdhury MS, Al-Mumin MA (2021) A systematic literature review on english and bangla topic modeling. *J Comput Sci* 17:1–18
- Alam MH, Rahoman MM, Azad MAK (2017) Sentiment analysis for Bangla sentences using convolutional neural network. In: 20th international conference of computer and information technology (ICCIT), New York, IEEE, pp 292–295
- Balahur A, Turchi M (2014) Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis. *Comput Speech Lang* 28:56–75
- Bengio Y, Lamblin P, Popovici D, Larochelle H (2006) Greedy layer-wise training of deep networks. In: Proceedings of the 19th international conference on neural information processing systems, MIT Press, Cambridge, pp 153–160
- Bishop CM (2006) Pattern recognition and machine learning 1 ed. Springer, New York
- Boccignone G, Bodini M, Cuculo V, Grossi G (2018) predictive sampling of facial expression dynamics driven by a latent action space. In: 2018 14th international conference on signal-image technology internet-based systems (SITIS), IEEE: New York, NY, USA, pp 143–150
- Bodini M (2019a) Aspect extraction from bangla reviews through stacked auto-encoders. *Data* 4:121
- Bodini M (2019b) Will the machine like your image? automatic assessment of beauty in images with machine learning techniques. *Inventions* 4:34
- Bodini M (2019c) A review of facial landmark extraction in 2D images and videos using deep learning. *Big Data Cogn Comput* 3:14
- Bodini M, D'Amelio A, Grossi G et al (2018) Single sample face recognition by sparse recovery of deep-learned LDA features. In: Blanc-Talon J, Helbert D, Philips W et al (eds) ACIVS 2018: advanced concepts for intelligent vision systems. Springer International Publishing, Cham, Switzerland, pp 297–308
- Burkov A (2019) The hundred-page machine learning book. Québec City, Canada, Andriy Burkov
- Chowdhury S, Chowdhury W (2014) Performing sentiment analysis in Bangla microblog posts. In: 2014 international conference on informatics, electronics & vision (ICIEV), IEEE: New York, NY, USA, pp 1–6
- Hacid H, Cellary W, Wang H, Paik HY, Zhou R (2018) Web information systems engineering—WISE 2018. Springer, Cham, Switzerland
- Harris ZS (1954) Distributional Structure. *WORD* 10:146–162
- Hasan KMA, Rahman M, Badiuzzaman S (2014) Sentiment detection from Bangla text using contextual valency analysis. In: 2014 17th international conference on computer and information technology (ICCIT), IEEE: New York, NY, USA, pp 292–295
- Hassan A, Amin MR, Al Azad AK, Mohammed N (2016) Sentiment analysis on bangla and romanized bangla text using deep recurrent models. In: International workshop on computational intelligence (IWCI), New York, IEEE, pp 292–295
- Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313:504–507
- Hinton GE, Osindero S, Teh YWA (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18:1527–1554
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence, Morgan Kaufmann Publishers Inc.: San Francisco, Vol 2, pp 1137–1143
- Lăzăroiu G, Neguriță O, Grecu I, Grecu G, Mitran PC (2020) Consumers' decision-making process on social commerce platforms: online trust, perceived risk, and purchase intentions. *Front Psychol* 11:890
- Li Z, Fan Y, Jiang B, Lei T, Liu W (2019) A survey on sentiment analysis and opinion mining for social multimedia. *Multimedia Tools Appl* 78:6939–6967
- Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space. In: 1st international conference on learning representations, ICLR 2013—workshop track proceedings. [arXiv:1301.3781](https://arxiv.org/abs/1301.3781): Ithaca, NY, USA
- Petz G, Karpowicz M, Fürschuß H, Auinger A, Střiteský V, Holzinger A (2015) Reprint of: computational approaches for mining user's opinions on the Web 2.0. *Inf Process Manag* 51:510–519
- Pontiki M, Galanis D, Papageorgiou H et al (2016) SemEval-2016 Task 5: aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), association for computational linguistics, Cambridge, pp 19–30
- Rahman MA, Dey EK (2018a) Aspect extraction from bangla reviews using convolutional neural network. In: 2018 Joint 7th international conference on informatics, electronics & Vision (ICIEV) and 2018 2nd international conference on imaging, vision & pattern recognition (icIVPR), IEEE: New York, NY, USA, pp 262–267
- Rahman MA, Dey EK (2018b) Datasets for aspect-based sentiment analysis in bangla and its baseline evaluation. *Data* 3:15
- Rifai S, Vincent P, Muller X, Glorot X, Bengio Y (2011) Contractive auto-encoders: explicit invariance during feature extraction. In: Proceedings of the 28th international conference on international conference on machine learning, Omnipress: Madison, WI, USA, pp 833–840
- Roy SS, Viswanatham VM, Krishna PV et al (2013) Applicability of rough set technique for data investigation and optimization of intrusion detection system. In: Awasthi AK (ed) Quality, reliability, security and robustness in heterogeneous networks Singh, K. Springer, Heidelberg, pp 479–484
- Roy SS, Mallik A, Gulati R et al (2017) A deep learning based artificial neural network approach for intrusion detection. In: Giri D, Mohapatra RN, Begehr H et al (eds) Mathematics and computing. Springer, Singapore, pp 44–53
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. *Nature* 323:533–536
- Salton G, Wong A, Yang CSA (1975) Vector space model for automatic indexing. *Commun ACM* 18:613–620
- Salton G, Yu CT (1973) On the construction of effective vocabularies for information retrieval. In: Proceedings of the 1973 meeting on programming languages and information retrieval, ACM, New York, pp 48–60
- Sazzed S, Sampath JA (2019) Sentiment Classification in Bengali and machine translated english corpus. In: 2019 IEEE 20th

- International conference on information reuse and integration for data science (IRI), IEEE: New York, NY, USA, pp 107–114
- Schütze H, Manning CD, Raghavan P (2008) Introduction to information retrieval. Cambridge University Press, Cambridge
- Sivic J, Zisserman A (2009) Efficient visual search of videos cast as text retrieval. *IEEE Trans Pattern Anal Mach Intell* 31:591–606
- Tang D, Wei F, Yang N, Zhou M, Liu T, Qin B (2014) Learning sentiment-specific word embedding for twitter sentiment classification. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, association for computational linguistics, Cambridge, vol 1, pp 1555–1565
- Trusov M, Bucklin RE, Pauwels K (2009) Effects of word-of-mouth versus traditional marketing: findings from an internet social networking site. *J Mark* 77:90–102
- Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th international conference on machine learning, ACM: New York, NY, USA, pp 1096–1103
- Wan X (2008) Using Bilingual Knowledge and ensemble techniques for unsupervised chinese sentiment analysis. In: Proceedings of the conference on empirical methods in natural language processing, association for computational linguistics: Cambridge, MA, USA, pp 553–561
- Weston J, Chopra S, Adams Kt (2014) #TagSpace: semantic embeddings from hashtags. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), association for computational linguistics, Cambridge, pp 1822–1827
- Yousefpour A, Ibrahim R, Hamed A (2014) A novel feature reduction method in sentiment analysis. *Int J Innov Comput* 4:34–40
- Yue L, Chen W, Li X, Zuo W, Yin M (2019) A survey of sentiment analysis in social media. *Knowl Inf Syst* 60:617–663
- Zhou X, Wan X, Xiao J (2016) Cross-lingual sentiment classification with bilingual document representation learning. In: Proceedings of the 54th annual meeting of the association for computational linguistics, association for computational linguistics, Cambridge, vol 1, pp 1403–1412

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.